

Prediksi Nilai Rating Aspek Film Berbasis Sentimen Analisis

Laporan Tugas Akhir

Diajukan Untuk Memenuhi

Persyaratan Guna Meraih Gelar Sarjana Strata 1

Teknik Informatika Universitas Muhammdiyah Malang



Oleh:

Harris Ishaq Ibrahim

201510370311056

Bidang Minat

Data Science

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMDIYAH MALANG

2020

LEMBAR PERSETUJUAN

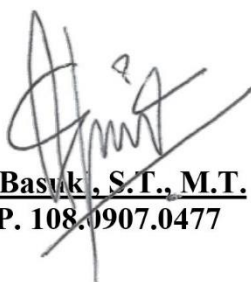
Prediksi Nilai Rating Aspek Film Berbasis Sentimen Analisis

TUGAS AKHIR

Sebagai persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

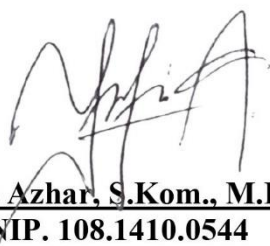
Menyetujui,
Malang, 13 Maret 2020

Dosen Pembimbing I



Setio Basuk, S.T., M.T.
NIP. 108.0907.0477

Dosen Pembimbing II



Yufis Azhar, S.Kom., M.Kom.
NIP. 108.1410.0544

LEMBAR PENGESAHAN

Prediksi Nilai Rating Aspek Film Berbasis Sentimen Analisis

TUGAS AKHIR

Sebagai persyaratan Guna Meraih Gelar Sarjana Strata I
Teknik Informatika Universitas Muhammadiyah Malang

Disusun Oleh:

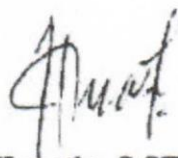
Harris Ishaq Ibrahim

201510370311056

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 1 bulan April tahun 2020

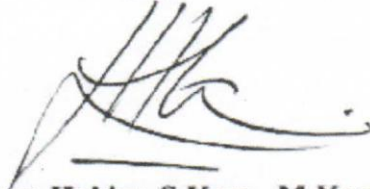
Menyetujui,

Penguji I



Nur Hayatin, S.ST., M. Kom.
NIP. 108.0907.0476

Penguji II



Luqman Hakim, S.Kom., M.Kom.
NIP. 108.1903.0658

Mengetahui,



a.n. Ketua Program Studi Informatika,

Gita Indan Marthasari, S.T., M.Kom.
NIP. 108.0611.0442

LEMBAR PERNYATAAN

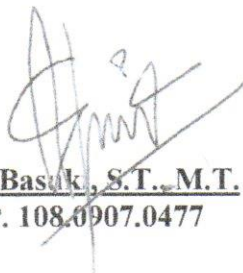
Yang bertanda tangan dibawah ini:

Nama : Harris Ishaq Ibrahim
NIM : 201510370311056
Fakultas / Jurusan : Teknik / Teknik Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Prediksi Nilai Rating Aspek Film Berbasis Sentimen Analisis”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun keseluruhan, kecuali dalam bentuk tujuan kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko / sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing


Setio Basuki, S.T., M.T.
NIP. 108.0907.0477

Malang, 13 Maret 2020

Yang Membuat Pernyataan



(Harris Ishaq Ibrahim)

LEMBAR PERSEMBAHAN

Alhamdulillah rabbil ‘alamin, dengan memanjatkan puji dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, dan tak lupa shalawat serta salam kepada junjungan Nabi Muhammad SAW, sehingga dengan ridha-Nya skripsi berjudul “Prediksi Nilai Rating Aspek Film Berbasis Sentimen Analisis” dapat terselesaikan.

Penyelesaian skripsi ini tidak lepas dari dukungan dan bantuan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis dengan segala kerendahan hati mengucapkan terimakasih dan penghargaan yang sebesar-besarnya kepada:

1. Allah SWT yang selalu memberikan kemudahan dan kesabaran dalam mengerjakan tugas akhir ini.
2. Orang Tua saya, Bapak Kismono dan Ibu Nana Sopiana, atas segala doa, dukungan, motivasi, dan nasehat yang diberikan selama mengerjakan tugas akhir ini.
3. Dosen Pembimbing saya, Bapak Setio Basuki, M.T. dan Bapak Yufis Azhar, M.Kom. yang sudah bersedia dan meluangkan waktunya untuk membantu dan membimbing terkait tugas akhir ini.
4. Seluruh Bapak dan Ibu Dosen Pengajar yang telah memberikan ilmunya dan Staff TU jurusan Teknik Informatika.
5. Sahabat – sahabat kelas B-GO yang tidak dapat saya sebutkan satu per satu. Terimakasih atas dukungan dan motivasi dari kalian.
6. Kawan Rantau Jatim Kos Notojoyo, sahabat seperjuangan dari SMK hingga perkuliahan ini yang telah banyak membantu selama ini.

Malang, 13 Maret 2020

Penulis

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul

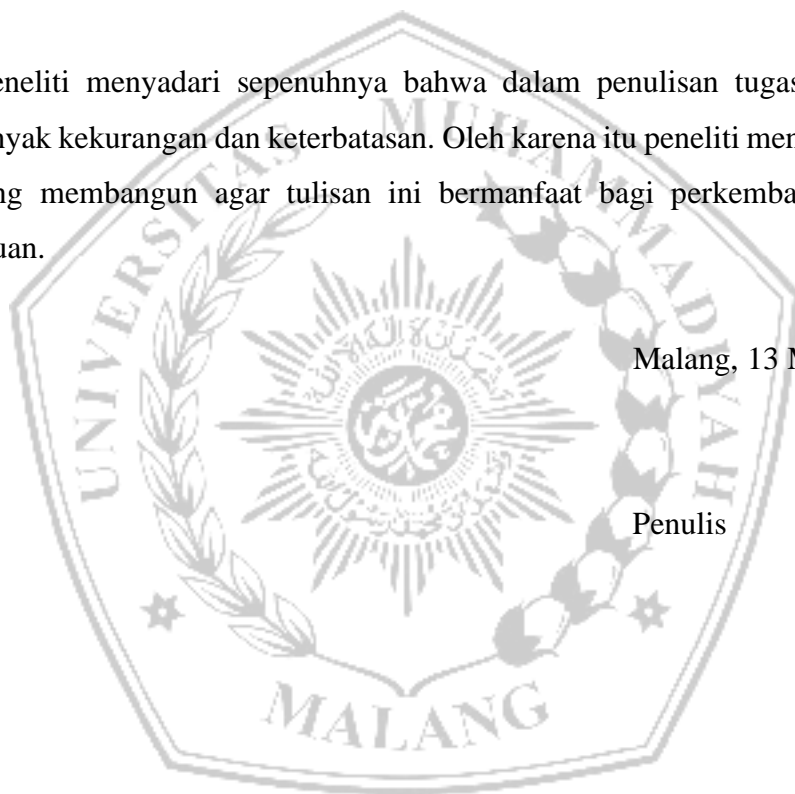
“PREDIKSI NILAI RATING ASPEK FILM BERBASIS SENTIMEN ANALISIS”

Di dalam tulisan ini disajikan pokok – pokok bahasan yang meliputi latar belakang, metode penelitian, dan hasil dan pembahasan yang telah didapat dari penelitian ini dan telah disimpulkan berdasarkan hasil yang telah didapat oleh peneliti.

Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 13 Maret 2020

Penulis



DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN	iii
ABSTRAK.....	iv
ABSTRACT.....	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 <i>Review Film</i>	6
2.3 <i>Text Mining</i>	7
2.4 <i>Data Preprocessing</i>	7
2.5 <i>Sentiment Analysis</i>	8
2.5.1 <i>Aspect – Based Sentiment Analysis</i>	8
2.6 <i>Word Embedding</i>	9
2.7 Convolutional Neural Network (CNN).....	9
2.8 Prediksi Rating	9
2.9 Evaluasi	10
2.9.1 <i>Accuracy</i>	10
2.9.2 <i>Precision</i>	10
2.9.3 <i>Recall</i>	10
BAB III METODOLOGI PENELITIAN	12
3.1 Pengumpulan Dataset.....	12
3.1.1 Jenis Dataset	12

3.1.2	Metode Pengumpulan Data	13
3.2	Text Preprocessing	15
3.3	Feature Engineering	17
3.4	Membuat Model Klasifikasi	18
3.5	Training Model	19
3.6	Testing Model	19
3.7	Evaluasi	19
3.7.1	Skenario Pengujian	19
3.8	Prediksi Rating	20
BAB IV HASIL DAN PEMBAHASAN.....		21
4.1	Kebutuhan Sistem	21
4.2	Dataset	21
4.3	Pembuatan Vocabulary	24
4.3.1	Memuat dataset	24
4.3.2	<i>Data Preprocessing</i>	26
4.3.3	<i>Filtering Minimum Occurance</i>	26
4.3.4	Menyimpan Hasil <i>Vocabulary</i>	27
4.4	Data Preprocessing	27
4.4.1	Pemberian Label	30
4.4	Embedding Data	31
4.5	Membuat Model CNN	32
4.5.1	Pelatihan dan Pengujian Model	34
4.6	Prediksi Review	36
4.6.1	Data Preprocessing	36
4.6.2	Klasifikasi Data	36
4.6.3	Prediksi Rating	36
4.7	Evaluasi	38
4.7.1	Perbandingan Model	39
BAB V KESIMPULAN.....		40
5.1	Kesimpulan	40
5.2	Saran	40
DAFTAR PUSTAKA		41
LAMPIRAN		43

DAFTAR GAMBAR

Gambar 1. Diagram Prosedur Penelitian	12
Gambar 2. Arsitektur CNN.....	18
Gambar 3. Detail Arsitektur CNN.....	34
Gambar 4. Grafik Perbandingan Nilai <i>Accuracy</i> Data Latih dan Data Uji	35
Gambar 5. Grafik Perbandingan Nilai <i>Loss</i> Data Latih dan Data Uji.....	35



DAFTAR TABEL

Tabel 1. Penelitian terdahulu	5
Tabel 2. List Kategori Aspek dan Label Class	13
Tabel 3. Contoh Data Review	14
Tabel 4. Tabel Hasil Proses <i>Data Preprocessing</i>	16
Tabel 5. Hasil Tahapan <i>Feature Engineering</i>	17
Tabel 6. Source Code Load Data Training	21
Tabel 7. Source Code Load Data Test.....	23
Tabel 8. Source Code Load Data	24
Tabel 9. Source Code Data Preprocessing.....	26
Tabel 10. Source Code Filtering Minimum Occurance	26
Tabel 11. Source Code Filtering Minimum Occurance	27
Tabel 12. Source Code Memuat Vocabulary	27
Tabel 13. Source Code Memuat Data Training	28
Tabel 14. Source Code Data Preprocessing.....	30
Tabel 15. <i>Source Code</i> Memberi Label Data	31
Tabel 16. Source Code Encoding Data	31
Tabel 17. Source Code Vocabulary Size	32
Tabel 18. <i>Source Code</i> Pembuatan Model CNN	32
Tabel 19. Source Code Fitting Model Network.....	34
Tabel 20. <i>Source Code</i> Train & Test Model.....	34
Tabel 21. Nilai <i>Accuracy</i> dan <i>Loss</i>	34
Tabel 22. <i>Source Code</i> Klasifikasi Data	36
Tabel 23. <i>Source Code</i> Memeriksa Hasil Klasifikasi	36
Tabel 24. <i>Source Code</i> Fungsi Menghitung Rating.....	37
Tabel 25. <i>Source Code</i> Menampilkan Hasil Prediksi Rating	38
Tabel 26. Hasil Prediksi Rating	38
Tabel 27. Nilai Precision dan Recall.....	38
Tabel 28. Perbandingan Nilai <i>Accuracy</i> dan <i>Loss</i>	39

DAFTAR PUSTAKA

- [1] D. Godes and D. Mayzlin, "Using online conversations to study word-of-mouth communication," *Mark. Sci.*, vol. 23, no. 4, 2004.
- [2] Anshuman, S. Rao, and M. Kakkar, "A rating approach based on sentiment analysis," *Proc. 7th Int. Conf. Conflu. 2017 Cloud Comput. Data Sci. Eng.*, no. January, pp. 557–562, 2017.
- [3] S. Budi, "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means," *Techno.Com*, vol. 16, no. 1, pp. 1–8, 2017.
- [4] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1733–1741, 2017.
- [5] C. Nanda, M. Dua, and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning," *Proc. 2018 IEEE Int. Conf. Commun. Signal Process. ICCSP 2018*, pp. 1069–1072, 2018.
- [6] S. Gojali, "Aspect Based Sentiment Analysis for Review Rating Prediction," *Int. Conf. Adv. Informatics Concepts, Theory Appl.*, 2016.
- [7] T. T. Thet, J. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," no. November, 2010.
- [8] V. Parkhe, "Aspect Based Sentiment Analysis of Movie Reviews," *Int. Conf. Soft Comput. Mach. Intell.*, 2014.
- [9] A. Rahman, "Online News Classification Using Multinomial Naive Bayes," *Itsmart*, vol. 6, no. 1, pp. 32–38, 2017.
- [10] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data Preprocessing for Supervised Learning," no. January, pp. 110–117, 2006.
- [11] G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," no. June, pp. 38–43, 2018.
- [12] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," vol. 2, no. 9, pp. 2909–2916, 2018.

- [13] T. M., Janyce Wiebet, Rebecca, "Development and Use of a Gold - Standard Data Set for Subjectivity Classifications," *Proc. 37th Annu. Meet. Assoc. Comput. Linguist. Comput. Linguist.*, p. 246, 1998.
- [14] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," *Determ. Sentim. Opin. / Proc. 20th Int. Conf. Comput. Linguist.*, pp. 1367-es, 2004.
- [15] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: An on-line lexical database," *Int. J. Lexicogr.*, vol. 3, no. 4, pp. 235–244, 1990.
- [16] P. Chaovalit and L. Thou, "Movie review mining: A comparison between supervised and unsupervised classification approaches," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 00, no. C, p. 112, 2005.
- [17] C. Practice, B. Ay, M. Talo, I. Hallac, and G. Aydin, "Evaluating deep learning models for sentiment classification," no. July, 2018.
- [18] Y. Kim, "Convolutional neural networks for sentence classification," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014.
- [19] G. Ganu, "Beyond the Stars : Improving Rating Predictions using Review Text Content," no. WebDB, pp. 1–6, 2009.
- [20] D. A. Fauziah *et al.*, "Klasifikasi Berita Politik Menggunakan Algoritma K-nearest Neighbor (Classification of Political News Content using K-Nearest Neighbor) Abstrak," *J. Sist. Inf. Univ. Jember*, vol. 6, no. 2, p. 8, 2018.
- [21] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," no. May, p. 4, 2018.
- [22] H. Wang and B. Raj, "A Survey: Time Travel in Deep Learning Space: An Introduction to Deep Learning Models and How Deep Learning Models Evolved from the Initial Ideas," no. October, 2015.

LAMPIRAN

Lampiran 1. Contoh data yang diprediksi.

Data Prediksi
<p>Birds of Prey memang panggungnya Margot Robbie. Harley Quinn is Margot Robbie and Margot Robbie is Harley Quinn, Margot Robbie disini seakan menjelaskan bahwa karakter Harley Quinn memang sejatinya harus diperankan oleh dirinya, mulai dari gestur tubuh, ekspresi wajah sampai cara bicara pun wajib mendapat 2 jempol.</p> <p>Memang Birds of Prey gk punya cerita yang berat untuk membuatnya terlihat sangat rumit dan gelap, tapi itulah yang bikin film ini likable, tidak berusaha keras untuk terlihat rumit dan gelap melainkan banyak diisi dengan momen - momen lucu.</p> <p>Scene peledakan Ace Chemical dan penyerbuan di kantor polisi merupakan scene terbaik di film ini, selain final fight terakhir melawan Black Mask ya walaupun gk spesial - spesial banget tapi tetap fun. Scene yang disebutkan sebelumnya menjadi sangat ikonik berkat kinerja tim production design membuatnya terlihat sangat keren dan ikonik, api ledakan pabrik kimia, kepulan asap warna-warni dan taburan konfeti, hingga tulisan neon yang dipakai memperkenalkan karakter - karakter secara playful membuatnya seakan mengikuti dunia Harley Quinn.</p> <p>Tentunya hal lain yang menjadi kunci adalah jajaran karakter menarik lainnya yang diperankan dengan baik juga oleh para pemerannya.</p> <p>And in the end, it's really worth to watch. EWWW I love it.</p>



UNIVERSITAS MUHAMMADIYAH MALANG
FAKULTAS TEKNIK
PROGRAM STUDI TEKNIK INFORMATIKA

Jl. Raya Tlogomas 246 Malang 65144 Telp. 0341 - 464318 Ext. 247, Fax. 0341 - 460782

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Harris Ishaq Ibrahim
NIM : 201510370311056
Judul TA : Prediksi Nilai Rating Aspek Film Berbasis Sentimen Analisis

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	8%
2.	Bab 2 – Daftar Pustaka	25 %	16%
3.	Bab 3 – Analisis dan Perancangan	25 %	6%
4.	Bab 4 – Implementasi dan Pengujian	15 %	4%
5.	Bab 5 – Kesimpulan dan Saran	5 %	5%
6.	Makalah Tugas Akhir	20%	4%

*) Hasil cek plagiarism diisi oleh salah satu pembimbing

*) Maksimal 5 kali

Mengetahui,
Dosen Pembimbing

(.....*Anam*.....)